

陕北黄土高原月降雨理论分布的初步研究

张少伟¹, 杨勤科^{2,1}, 陆广勇¹

(1. 中国科学院 水利部 水土保持研究所, 陕西 杨凌 712100; 2. 西北大学 城市与环境学院, 西安 710069)

摘 要: 利用陕北两个典型气象站点多年降雨强度数据, 使用数学方法描述月雨强分布的形状, 从而满足水文模型、土壤侵蚀模型等对降雨强度数据的需求。通过探讨小时雨强在月时间尺度上应符合的理论分布函数, 并采用统计学参数估计方法进行参数拟合, 来选择合适的数学方法对月雨强分布进行描述。使用最大似然法对常见的 7 种理论分布函数进行拟合的结果表明, LOGN、GAMA、WBL 等理论分布函数都可以较好地再现月降雨强度分布, 90% 以上的月份 COE 值大于 0.7, 一半以上的月份 COE 值大于 0.9。另外, WBL 分布对降雨强度峰值亦有良好的表现能力, 因此推荐在陕北黄土高原降雨分布的模拟过程中使用。

关键词: 月降雨强度分布; 理论分布函数; 最大似然法

中图分类号: P426.61+4

文献标识码: A

文章编号: 1005-3409(2011)03-0001-04

The Theoretical Distribution of Monthly Rainfall in Northern Shaanxi Loess Plateau

ZHANG Shao-wei¹, YANG Qin-ke^{2,1}, LU Guang-yong¹

(1. Institute of Soil and Water Conservation, Chinese Academy of Sciences and Ministry of Water Resources, Yangling Shaanxi, 712100, China; 2. College of Urban and Environmental Science, Northwest University, Xi'an 710069, China)

Abstract: Based on the hourly datasets of rainfall intensity from two typically climate data stations in northern Shaanxi, the paper studied the optimal theoretical distribution function of hourly rainfall intensity on the monthly temporal scale in this area. The aim is to improve the performance of hydrological and soil conservation models. When rainfall intensity datasets are used, the models will be greatly improved. The results showed that parameters of TDFs (including LOGN, GAMA and WBL etc.), estimated by maximum likelihood method, can reproduce the shape of monthly distribution of rainfall intensity very well. Values of COE were greater than 0.7 for more than 90% of months and greater than 0.9 for more than 50% of months. Meanwhile, the distribution of WBL is recommended in study of rainfall distribution in northern Shaanxi due to its good performance in reproducing the peak value of rainfall intensity.

Key words: monthly rainfall distribution; theoretical distribution function; maximum likelihood method

在水文径流、土壤侵蚀和污染迁移模型预测中使用高时间分辨率的降雨强度数据可以大幅提高模型的动态预测能力^[1-2], 然而我国大多数气象站点仅提供日降雨观测数据, 时间间隔更小的雨强数据难以获得, 因此如何使用现有的气象观测数据来估计出应用中适用的雨强数据就显得十分重要。Kandel 等指出使用 2 min 间隔的雨强观测数据进行土壤侵蚀模拟的有效系数为 0.8, 而使用日降雨则仅仅为 0.26, 直接使用日降雨数据的模拟结果不佳是由于时间的平滑作用^[3]。由于连续观测的雨强数据难以获得, 雨强

随时间分布研究作为一种改善降雨数据质量的折中方式应运而生^[4], 它通过使用现有的理论统计分布函数对雨强的时间分布进行描述, 获得特定地区雨强统计分布曲线。雨强分布研究起源于降雨分解研究, 通过对雨强统计分布的随机分解实现了大量随机天气发生器, 另外雨强分布也可以通过转化直接应用于预测模型^[3]。国内研究主要集中于降雨强度的应用方面, 典型研究有谢云利用雨强进行降雨侵蚀力和降雨特征及其区域分异的研究^[5-6], 而如何表现雨强理论分布形态的研究鲜有发表。本文将以陕北地区两个

收稿日期: 2011-02-24

修回日期: 2011-03-14

资助项目: 国家重点基础研究发展计划项目专题: 区域土壤侵蚀模型研究(2007CB407204-03); 黄土高原水土流失数学模型(第一期)研发(黄水保 200651-08)

作者简介: 张少伟(1985-), 男, 陕西省户县人, 研究生在读, 主要从事基于 GIS 的区域水土流失研究。Email: swei_117@yahoo.com.cn

通信作者: 杨勤科(1962-), 男, 陕西省陇县人, 教授, 博士, 主要从事基于 GIS 的区域水土流失研究。E-mail: qkyang@nwu.edu.cn

典型气象站点长时期降雨强度累计资料为基础数据,探讨该区域小时雨强在月时间尺度上的理论分布方式,本研究对于改进和完善区域土壤侵蚀定量评价技术具有重要意义。

1 研究方法

1.1 数据来源

本文选择绥德站和延安站分别代表黄土丘陵 I 副区和黄土丘陵 II 副区进行研究。该区属于中温带干旱大陆性季风气候,降雨少且不匀,夏季多雨,冬季受干燥少雨,降雨主要集中在 5- 9 月时段,4 月和 11 月只有少数年份有降雨,其它月份几乎没有降雨。本文使用的降雨观测数据为延安、绥德两气象站点的逐日小时降雨观测资料,来自陕西省降水强度数据集,每日的观测从 20:00 开始,1961- 2000 年共 40 a 的连续观测。

为了能够进行月雨强样本的数值计算,选取 3 个条件对数据进行筛选: (1) 整个月份的降雨记录完整无缺; (2) 参照 Anderson 在日雨强分布的研究,本文选择小时雨强大于阈值(1.0 mm/ h) 的降雨用于月雨强分布估计,这样可以消除较小的雨强数据观测本身的误差,另外水文和土壤侵蚀研究中也更关心较大的降雨强度; (3) 由于参数估计方法的数学限制,月内降雨的小时数必须不少于 4 个^[7]。根据上述条件进行数据筛选后,获得可用于月降雨强度分布拟合的月份样本,延安站样本 226 个,绥德站 190 个。

1.2 理论分布函数

本文将使用 7 种不同的理论分布函数对月雨强分布进行描述,并评价其适合程度,7 种理论分布函数的简称、全称见表 1,各种分布函数需要的参数个数在括号中标出。指数分布仅有一个参数,通用极值分布参数为 3 个,其他分布均为两个参数,带有星号的理论分布函数为 3 种极值分布类型。陕北地区降雨常出现短时暴雨,强降雨对土壤侵蚀的作用巨大,而极值分布常用于描述数据中的极端值,因此特别将 3 种极值分布包含在内。

表 1 理论分布函数

理论分布函数	简称	名称
Exponential	EXP(1)	指数分布
LogNormal	LOGN(2)	对数正态分布
Gamma	GAMA(2)	伽马分布
Generalized Pareto	GP(2)	帕累托分布
* Generalized Extreme Value	GEV(3)	通用极值分布
* Weibull	WBL(2)	威尔布分布
* Gumbel	GUMB(2)	Gumbel 极值分布

1.3 参数估计方法

目前进行参数估计的方法有矩估计法、最小二乘法和最大似然法等。最大似然法利用样本的联合密度函数达到最大进行参数估计,与理论分布函数的形式联系更加紧密。前两种方法在降雨分布研究中有应用,如 Anderson 等在澳大利亚日降雨分布研究中使用了矩估计法进行参数拟合^[4],Ming Li 在尼泊尔的研究中采用了最小二乘法^[8],但一般而言矩估计法多用于大样本的参数估计,最小二乘法则多用于线性回归的参数估计,因此本文采用的最大似然法进行参数估计,其进行参数估计的根据是最大似然函数值最大。

1.4 理论降雨强度序列

经过参数估计并带入理论分布函数的数学形式,即可得到完整的月降雨强度理论分布函数,该函数有概率密度函数和累计分布函数两种形式。累计分布函数形式用来描述小于变量的值出现的概率之和,其反函数被称为反累计分布函数。将降雨量总和作为约束条件,在理论分布函数上按一定规则进行采样即得到理论估计的降雨强度。本文重点在于探讨适用于陕北月雨强分布的理论函数,采样规则利用观测雨强序列的超越概率(Probability Of Exceedance), 首先将月降雨强度序列进行排序,对应有效降雨强度($P_1 \leq P_2 \leq \dots \leq P_i$), 第 i 个雨强在理论分布函数上对应超越概率 x_i , Cunnane 的研究中将超越概率的计算方法定义为^[9]:

$$x_i = (i - 0.4) / (m + 0.2) \tag{1}$$

式中: m ——有效降雨强度数值的个数,将超越概率 x_i 代入反累积分布函数即可得到对应的估计雨强值,估计雨强集合即为理论降雨强度序列。

1.5 评价方法

本文对各种理论分布函数与降雨强度分布之间吻合度的评价有两个方面,一是选择有效系数作为定量标准评价理论分布和观测月降雨之间的吻合程度,将样本个数 10% 和 50% 分位数上的有效系数作为标志。有效系数 COE(Coefficient of Efficiency) 用于评价两个数据序列的差异大小,本文中用于表现月样本的观测降雨强度序列(P_i) 和估计降雨强度序列(\hat{P}_i) 之间差异的大小,COE 计算方法为^[10]:

$$COE = 1 - \frac{\sum_{i=1}^m (P_i - \hat{P}_i)^2}{\sum_{i=1}^m (P_i - \overline{P_i})^2} \tag{2}$$

式中: $\overline{P_i}$ ——观测降雨强度的均值; m ——当月雨强序列中小时雨强的个数,每个月降雨样本可计算出一个有效系数。评价的另一面是检测理论分布函数对月降雨强度最大值的描述能力,用理论分布函数计算出的理论雨强峰值和观测雨强峰值计算有效系数和决定

系数, 决定系数代表观测值和理论值之间相关性的
大小, 有效系数代表直接用理论值代表观测值的
有效性。

2 结果与分析

对延安、绥德两站小时降雨数据进行整理, 获
得两站的月降雨的小时记录, 以月样本作为研究
对象进行月雨强分布研究。陕北黄土高原地区冬
季降雨稀少, 40 a 中符合数据筛选要求的月份主
要集中在 5-9 月, 绥德这 5 个月的平均降雨小时
数为: 9.7, 13.9, 21.3, 22.8, 21.7 h, 延安站 5
个月的平均降雨小时数略大于绥德, 分别是 14.2,
16.2, 24.9, 27.0, 23.8 h。利用月降雨样本结
合最大似然法进行参数估计, 分别完成 7 种理
论分布函数的参数拟合。

2.1 理论分布函数对月降雨强度分布的模拟

将理论降雨强度序列与月降雨观测数据进行
比较, 以观测数据作为标准计算有效系数 COE, 从
而评价各种理论分布函数对月降雨分布的描述和
重现能力。当月内的有效降雨小于 10 h 时, GEV
分布估计的降雨出现极大的偏差, 其它 6 种理
论分布函数的均方根误差如图 1 所示, 可见
LOGN、GAMA 和 WBL 三种函数较好地再现了
月降雨强度分布。

图 2 为 LOGN、GAMA 和 WBL 在延安站 COE
值的分布, 图中标出了样本个数 50% 和 10% 分
位数的位置。这 3 种理论分布与实测月降雨吻
合最好, 延安站分别有 50% 的月份 COE 值大于
0.90, 0.89, 0.89,

同时仅有 10% 的月份 COE 小于 0.73, 0.72, 0.74;
绥德站有 50% 的月份 COE 值大于 0.88, 0.86,
0.87, 同时仅有 10% 的样本 COE 小于 0.73,
0.73, 0.75。可见 3 种理论分布函数对月雨强
的整体描述能力相当, 延安和绥德两站的情况一
致。

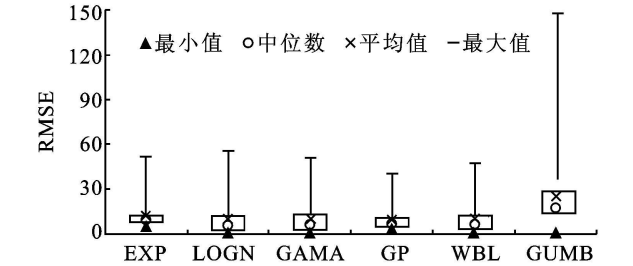


图 1 理论分布函数的 RMSE

其他几种理论分布函数中, EXP 分布为单一参
数的理论分布函数, 对月降雨分布的表达能力不
佳, 两站均有 20% 左右的月份 COE 值小于 0, 因
此认为 EXP 分布不适用雨强分布统计。GP 和
GUMB 分布的 COE 值优于 EXP 分布, 但仍出现
部分负数。值得注意的是, COE 值负数的月份有
效降雨的小时数较少, 这种情况下不利于理论分
布函数的参数估计, 可能是导致 COE 值负数的原
因。GEV 分布的 COE 值中出现个别非常大的负
值(E5 数量级以上), 该分布为三参数理论分布
函数, 除了均值和方差外还有一个形状参数对其
性质起决定作用, 忽略对形状参数的约束导致
该分布直接应用于月雨强分布时出现个别极端
值。

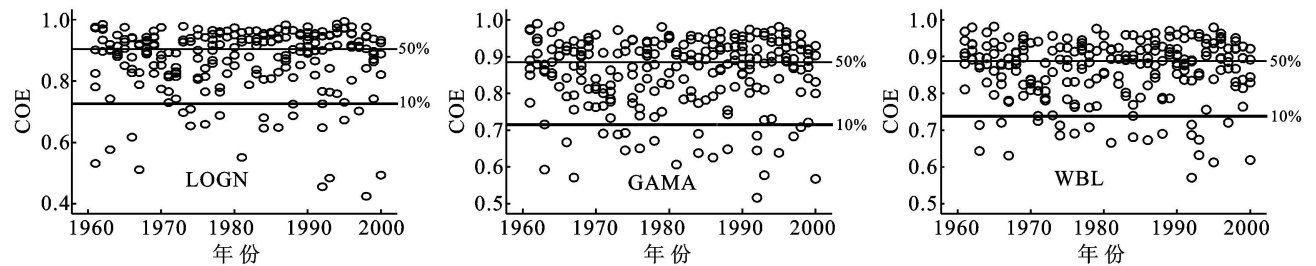


图 2 延安站不同理论函数分布的 COE

2.2 理论分布函数对月最大降雨强度的模拟

降雨强度峰值是水文、土壤侵蚀等研究中需
要考虑的重要因素^[1], 因此本文不仅考虑了理论
分布函数对月降雨整体的模拟能力, 同时将降雨
强度峰值在理论分布函数上的表达和再现作为另
一评价标准。文中降雨强度峰值即月雨强序列
的最大值, 根据月降雨强度峰值对应的超越概率,
可以计算累积分布函数对应的理论雨强峰值, 将
各月降雨强度峰值和理论峰值绘制散点图, 并计
算有效系数 COE 和决定系数 R^2 , 决定系数代
表观测值和理论值之间相关性的, 有效系数代
表直接用理论值表示观测值的有效性。

GEV 分布的有效系数 COE 和决定系数 R^2 出
现异常, 其他 6 种理论分布的 COE 值和 R^2 见
表 2, 可

见极值分布对月雨强峰值的模拟相当好, 特别
是 GUMB 分布在两个站点的雨强峰值 COE 值均
达到 0.85, 决定系数达到 0.98, 对雨强极值模
拟的适用性非常好。GP 分布是另一种能有效表
现雨强极值的理论分布函数, 但值得注意的是
WBL 分布虽然不及上述两种理论分布对表现好,
但对月雨强分布整体的模拟也有良好的效果, 是
一种适用于陕北月降雨分布模拟的理论函数。

由图 3 可以看出, GP、WBL、GUMB 三种理
论分布函数对雨强极值的模拟效果良好, COE 值
均超过 0.7, 表现了直接用理论值来表达观测雨
强极值是可行的。但理论雨强峰值仍倾向于低估
雨强峰值, 多数的研究结果中也有类似的结论,
即人为模拟极端降雨

时偏向于低估其值, 此三种理论分布函数多年平均的低估值依次为 11%、17% 和 15%。值得注意的是理论值和观测值之间有很强的线性关系, WBL 分布决定系数值超过 0.85, GP、GUMB 分布超过 0.95, 因此有可能在特定站点用理论雨强峰值的线性变换来进一步提高对观测值的表达精度。

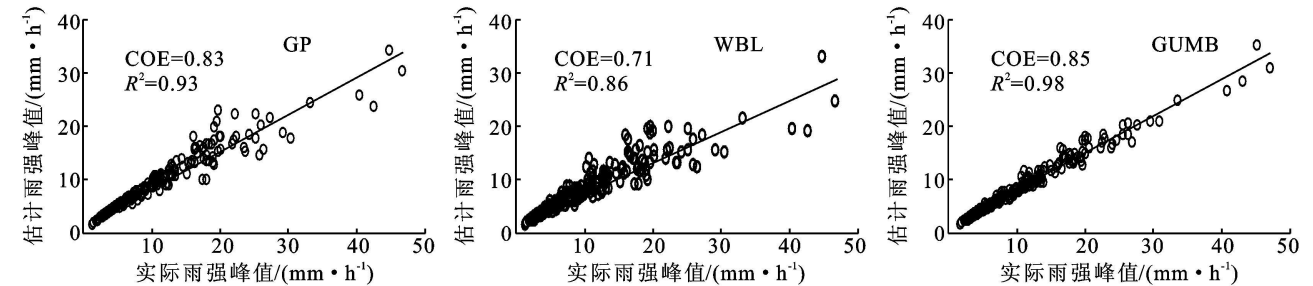


图 3 延安站不同理论函数分布对最大降雨强度的估计

3 结论

(1) 利用历史降雨强度数据, 通过最大似然法进行参数拟合, LOGN、GAMA 和 WBL 三种理论分布函数能够较好地再现月降雨强度分布。三种理论分布函数对月雨强的表现能力相当, LOGN 分布略优于其它两种分布。

(2) 降雨强度峰值是水文、土壤侵蚀研究中的重要参数, GP、WBL、GUMB 三种理论分布函数较好地再现了月降雨强度峰值, 极值分布 GUMB 明显强于其它理论分布函数。理论分布对降雨强度峰值有一定的低估, 低估的范围约 15% 左右, 理论模拟的雨强峰值和观测值有很好的线性关系。

(3) 对陕北小时雨强在月时间尺度上的分布来说, WBL 是一种有用的理论分布函数, 虽然其对月雨强分布和月雨强峰值的表现均不是最好, 但作为一种极值分布, 它同时对二者有较好的再现。在陕北气候降雨条件下对月降雨强度分布进行模拟时, WBL 分布将是一种行之有效的理论分布函数。

通过月降雨强度理论分布研究, 表明将 WBL 分布应用于陕北黄土高原两个典型气象站点月雨强分布的表达有明显优势, 但在整个黄土高原上是否适用仍需要更多典型站点资料进行验证。另外通用极值分布 GEV 是一种被广泛应用的理论分布函数, 但本研究中未能成功对其形状参数进行约束, 因此在研究中加入适用于月降雨分布的约束条件十分重要。最后, Kandel 等研究中应用了 LOGN 分布来表现日雨强分布, 但本研究中该分布表现月雨强极值的效果不佳, 在陕北的日雨强分布是否符合 LOGN 分布也有待进一步研究。

表 2 雨强峰值有效系数 COE 与决定系数 R ²							
观测站	指标	EXP	LOGN	GAMA	GP	WBL	GUMB
延安	COE	0.66	0.64	0.65	0.83	0.71	0.85
	R ²	0.72	0.69	0.81	0.93	0.86	0.98
绥德	COE	0.65	0.61	0.64	0.79	0.69	0.85
	R ²	0.73	0.66	0.80	0.87	0.84	0.98

参考文献:

[1] Koutsoyiannis D. Rainfall disaggregation methods: theory and applications[C]. Workshop on Statistical and Mathematical Methods for Hydrological Analysis, 2003.

[2] Zhang X, Garbrecht J. Evaluation of Cligen Precipitation Parameters and Their Implication on Wepp Runoff and Erosion Prediction[J]. Transactions of the ASAE, 2002, 46(2) : 311-320.

[3] Kandel D, Western A, Grayson R, et al. Process parameterization and temporal scaling in surface runoff and erosion modelling[J]. Hydrological Processes, 2004, 18(8) : 1423-1446.

[4] Anderson B, Siriwardena L, Western A, et al. Which theoretical distribution function best fits measured within day rainfall distributions across Australia[C]. 30th Hydrology and Water Resources Symposium, 2006.

[5] 谢云, 林小鹃, 殷水清, 等. 中国次雨量特征及其区域分异[J]. 地理学报, 2009, 64(3) : 323-330.

[6] 谢云, 章文波. 用日雨量和雨强计算降雨侵蚀力[J]. 水土保持通报, 2001, 21(6) : 53-56.

[7] Kirby W. Algebraic boundedness of sample statistics [J]. Water Resources Research, 1974, 10(2) : 220-222.

[8] Li M, Shao Q, Renzullo L. Estimation and spatial interpolation of rainfall intensity distribution from the effective rate of precipitation[J]. Stochastic Environmental Research and Risk Assessment, 2010, 24(1) : 117-130.

[9] Cunnane C. Unbiased plotting positions: a review [J]. Journal of Hydrology, 1978, 37(3/ 4) : 205-222.

[10] Legates D, McCabe G Jr. Evaluating the use of ‘ goodness-of-fit’ measures in hydrologic and hydroclimatic model validation[J]. Water Resources Research, 1999, 35(1) : 233-241.

[11] Parsons A, Stone P. Effects of intra-storm variations in rainfall intensity on interrill runoff and erosion[J]. Catena, 2006, 67(1) : 68-78.